

```

1  ATGCCCTTTT GCCACAATAT AATTAATATT TCCTGTGTGA AAAACAACCTG
51  GTCAAATGAT GTCCGTGCTT CCCTGTACAG TTTAATGGTG CTCATAATTC
101 TGACCACACT CGTTGGCAAT CTGATAGTTA TTGTTTCTAT ATCACACTTC
151 AAACAACCTC ATACCCCAAC AAATTGGCTC ATTCATTCCA TGGCCACTGT
201 GGACTTTCTT CTGGGGTGTC TGGTCATGCC TTACAGTATG GTGAGATCTG
251 CTGAGCACTG TTGGTATTTT GGAGAAGTCT TCTGTAAAAT TCACACAAGC
301 ACCGACATTA TGCTGAGCTC AGCCTCCATT TTCCATTTGT CTTTCATCTC
351 CATTGACCGC TACTATGCTG TGTGTGATCC ACTGAGATAT AAAGCCAAGA
401 TGAATATCTT GGTATTTTGT GTGATGATCT TCATTAGTTG GAGTGTCCCT
451 GCTGTTTTTG CATTTGGAAT GATCTTTCTG GAGCTAAACT TCAAAGCGC
501 TGAAGAGATA TATTACAAAC ATGTTCACTG CAGAGGAGGT TGCTCTGTCT
551 CTTTTCAGCA AATATCTGGG GTACTGACCT TTATGACTTC TTTTATATA
601 CCTGGATCTA TTATGTTATG TGTCTATTAC AGAATATATC TTATCGCTAA
651 AGAACAGGCA AGATTAATTA GTGATGCCAA TCAGAAGCTC CAAATTGGAT
701 TGGAAATGAA AAATGGAATT TCACAAAGCA AAGAAAGGAA AGCTGTGAAG
751 ACATTGGGGA TTGTGATGGG AGTTTTCCTA ATATGCTGGT GCCCTTTCTT
801 TATCTGTACA GTCATGGACC CTTTCTTCA CTACATTATT CCACCTACTT
851 TGAATGATGT ATTGATTTGG TTTGGCTACT TGAACCTAC ATTTAATCCA
901 ATGGTTTATG CATTTTTCTA TCCTTGGTTT AGAAAAGCAC TGAAGATGAT
951 GCTGTTTGGT AAAATTTTCC AAAAAGATTC ATCCAGGTGT AAATTATTTT
1001 TGAATTGAG TTCATAG

```

FEATURES:

Start: 1
Stop: 1015

HOMOLOGOUS PROTEIN:

Top 10 BLAST Hits:

gi 7657142 ref NP_055441.1	G protein-coupled receptor 58 >gi 6...	317	8e-86
gi 7657140 ref NP_055442.1	G protein-coupled receptor 57 >gi 6...	316	2e-85
gi 4505925 ref NP_003958.1	putative neurotransmitter receptor ...	264	9e-70
gi 3941547 gb AAC82381.1	(AF069546) putative odorant receptor ...	226	2e-58
gi 4028153 gb AAC96117.1	(AF083221) putative neurotransmitter ...	210	1e-53
gi 6224984 sp O70528 5H4_CAVPO	5-HYDROXYTRYPTAMINE 4 RECEPTOR (...)	209	4e-53
gi 3646424 emb CAA09599.1	(AJ011370) serotonin 4 receptor [Rat...	208	7e-53
gi 3326989 emb CAA73108.1	(Y12506) 5-HT4 receptor [Homo sapiens]	206	3e-52
gi 3326991 emb CAA73109.1	(Y12507) 5-HT4 receptor [Homo sapiens]	206	3e-52
gi 3646278 emb CAA09600.1	(AJ011371) serotonin 4 receptor [Hom...	206	3e-52

blast to dbEST:

gi|7906115|gb|AW813121.1|AW813121 CM4-ST0189-051099-021-h04 ST0... 436 e-120

Tissue expression for modulatory use:

EST Expression:

gi|7906115|gb|AW813121.1|AW813121 : Stomach

cDNA panel screen:

Human Brain
Human fetal Brain
Human placenta
Human liver
Human kidney

FIGURE 1

```

1  MPFCHNIINI SCVKNNWSND VRASLYSLMV LIILTTLVGN LIVIVSISHF
51 KQLHTPTNWL IHSMATVDFL LGCLVMPYSM VRSAEHCWYF GEVFCKIHTS
101 TDIMLSSASI FHLSFISIDR YYAVCDPLRY KAKMNILVIC VMIFISWSVP
151 AVFAFGMIFL ELNFKGAEEI YYKHVHCRGG CSVFFSKIIG VLTFTSFYI
201 PGSIMLCVYY RIYLIAKEQA RLISDANQKL QIGLEMKNGI SQSKERKAVK
251 TLGIVMGVFL ICWCPFFICT VMDPFLHYII PPTLNDVLIW FGYLNSTFNP
301 MVYAFFYPWF RKALKMMLFG KIFQKDSSRC KLFLELSS

```

FEATURES:

Functional domains and key regions:

[1] PDOC00001 PS00001 **ASN_GLYCOSYLATION**

N-glycosylation site

Number of matches: 3

```

1      9-12 NISC
2     16-19 NWSN
3    295-298 NSTF

```

[2] PDOC00005 PS00005 **PKC PHOSPHO SITE**

Protein kinase C phosphorylation site

327-329 SSR

[3] PDOC00006 PS00006 **CK2 PHOSPHO SITE**

Casein kinase II phosphorylation site

Number of matches: 3

```

1     99-102 TSTD
2    270-273 TVMD
3    283-286 TLND

```

[4] PDOC00007 PS00007 **TYR PHOSPHO SITE**

Tyrosine kinase phosphorylation site

Number of matches: 2

```

1     82-89 RSAEHCWY
2    165-172 KGAEIYY

```

[5] PDOC00008 PS00008 **MYRISTYL**

N-myristoylation site

Number of matches: 2

```

1    239-244 GISQSK
2    253-258 GIVMGV

```

[6] PDOC00210 PS00237 **G PROTEIN RECEPTOR**

G-protein coupled receptors signature

108-124 ASIFHLSFISIDRYAV

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	27	47	1.997	Certain
2	61	81	1.301	Certain
3	99	119	1.200	Certain
4	136	156	2.270	Certain
5	189	209	1.960	Certain
6	252	272	2.448	Certain
7	287	307	1.057	Certain

BLAST Alignment to Top Hit:

>gi|7657142|ref|NP_055441.1| G protein-coupled receptor 58
>gi|6739494|gb|AAF27278.1|AF112460_1 (AF112460) G
protein-coupled receptor 58 [Homo sapiens]
Length = 306

Score = 317 bits (805), Expect = 8e-86

Identities = 155/301 (51%), Positives = 202/301 (66%), Gaps = 13/301 (4%)

Query: 26 LYSLMVLIIILTTLVGNLIVIVSISHFKQLHTPTNWLHSMATVDLLGCLVMPYSMVRS 85
+YS M I T+ GNL +I+SI+FKQLHTPTN+LI SMA DFLG +MPYSM+RS
Sbjct: 1 MYSFMAGSIFITIFGNLAMIISISYFKQLHTPTNFLILSMAITDFLGFTIMPYSMIRSV 60

Query: 86 EHCWYFGEVFCKIHTSTDIMLSSASIFHLSFISIDRYAVCDPLRYKAKMNILVICVMIF 145
E+CWYFG FCKI+ S D+MLS SIFHL ++IDR+YA+C PL Y K+ I VI ++
Sbjct: 61 ENCWYFGLTFCKIYYSFDLMLSITSIFHLCSVAIDRFYAICYPLLYSTKITIPVIKRLLL 120

Query: 146 ISWSVPAVFAFGMIFLELNFKGAEEIYKHHVHCRGGCSVFFSKISGVLTFMTSFYIPGSI 205
+ WSVP FAFG +F E G E Y V C C V F+K+ G FM F+ PGS+
Sbjct: 121 LCWSVPGAFAFGAVFSEAYADGIEG-YDILVACSSSCPVMFNKLWGTTLFMAGFFTPGSM 179

Query: 206 MLCVYYRIYLIAKEQARLISD--ANQKLQIGLEMKNGISQSKERKAVKTLGIVMGVFLIC 263
M+ +Y +I+ ++++ A I++ NQ Q+ K++KA KTLGIV+GVFL+C
Sbjct: 180 MVGIYGKIFAVSRKHAHAINNLRENQNNQV-----KKDKKAAKTLGIVIGVFLLC 229

Query: 264 WCPFFICTVMDPFLHYIIPPTLNDVLIWFGYLNSTFNPMVYAFFYPWFRKALKMMLFGKI 323
W P F ++DPFL++ P L D L WFGY NST NP++Y FFYPWFR+ALK +L GKI
Sbjct: 230 WFPCFFTILLDPFLNFSTPVVLFDAITWFGYFNSTCNPLIYGFFYPWFRRALKYILLGKI 289

Query: 324 F 324
F
Sbjct: 290 F 290

Hmmer search results (Pfam):

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
CE00516	CE00516 SEROTONIN_RECEPTOR	208.8	8.5e-59	3
PF00001	7 transmembrane receptor (rhodop	203.1	2.9e-63	1
CE00507	CE00507 DOPAMINE_RECEPTOR	154.5	1.2e-42	3
CE00267	E00267 Octopamine receptor	133.5	6.6e-37	3
CE00503	CE00503 ADRENERGIC_RECEPTOR	128.1	8e-35	3
CE00510	CE00510 HISTAMINE_RECEPTOR	124.6	1.7e-35	3
CE00323	E00323 kappa_opioid_receptor	49.3	2.2e-14	2
CE00336	E00336 orexin_receptor	47.3	7.8e-13	2
CE00326	E00326 mu_opioid_receptor	43.2	5.1e-12	2
CE00177	CE00177 ADENOSINE	39.4	3.8e-10	2
CE00205	CE00205 GH_secretagogue	37.0	1.6e-12	1
CE00328	E00328 tachykinin_receptor	33.1	1.9e-08	2
CE00513	CE00513 NEUROPEPTIDE_Y_RECEPTOR	24.9	1.8e-06	1
CE00530	CE00530 CHEMOKINE_RECEPTOR_TYPE	24.0	5.2e-07	2
CE00266	CE00266 Conopressin_receptor	19.5	0.00016	1

1	TGTGTGCCTT	CCCCACGGCC	CACATGGAAA	CTCAGCTTCC	CTGGGCTCTC
51	ACCTCAGTGC	TCTTTGCACC	ATGCCACTTC	ACTTCTTACC	ATAGACTTTA
101	GTCTGTATAT	CAAAAATCCT	TAATAAGTGG	GACCACAATA	TCATTCTTCA
151	CTAGTAGGTG	CCATTATTTA	AGACATCAAA	AATTCTTTATA	TGAAAATTTA
201	AACCTTAATT	TTTAAAAATT	GACCTCAAGT	TGGTTCTTTC	TGCAATAGAG
251	AGAAAGACTG	GACTGATATC	AAAGAAACTG	AACTAGTAAT	TGAGACCACA
301	GCTCATATTC	AAGCTTTAAT	GATTTGTGAA	TTAAACTTCA	ATTTCTTGCC
351	AGTAAAATGA	GGTTAATGAA	GCATGTCATG	TCTTCCTTAC	AAAAAATGAT
401	GTTGGTCTTA	TAAATGTCAA	ATAAACTGT	AACTGAAATA	ATGCATGGAA
451	TATATGTAGA	AGTACTTTGC	AAGCTGCAAA	TAAATATATG	TAAGCTTCTT
501	CTTATTACAA	TTAATTATAA	AATAAGAAAT	ATATAATTTA	CGTATGACCA
551	AGTGCAGAAG	TGGGCATAGT	GCGTATTAAT	TCTGCACAAG	AGGCACATTT
601	AAAAGACTTC	TTGTCTTTTT	CATAAACATA	TTTTAATTCC	AAATTTTCTG
651	GCAATTTTCT	TTCTACCTTT	TACAATGTAA	GTCTCCCCCA	CTAGACTGGC
701	TTCTATGGA	CACAGTATTG	TTTCCATAGA	TTAAGATGTG	TGATCATAAT
751	CATCATTTGA	TCTTCTCTGA	CCTACAAATG	TGGCAGTTTC	ATATAGTTCA
801	ACCTAATAAT	TTAGCTTCAT	ATTTCTAGAA	GAATGTCCAG	CTCCAGGTAA
851	AAGTTTTTTA	AAAATTGCTG	TACTGAACATA	TTGAATGGAA	CTTGGAATA
901	AAGTCCCTTC	CAAAATAACT	ATTCTTCAAC	AGAGAGTAAT	AGGTAAATGT
951	TTTAGAAGTG	AGAGGACTCA	AATTGCCAAT	GATTTACTCT	TTTATTTTTC
1001	CTCCTAGGTT	TCTGGGATAA	GTATGTGCAA	ATAAAAAATA	AACATGAGAA
1051	GGAACGTGTA	CCTGATTATG	GATTTGGGAA	AAAGATAAAT	CAACACACAA
1101	AGGGAAAAGT	AAACTGATTG	ACAGCCCTCA	GGAATGATGC	CCTTTTGCCA
1151	CAATATAAAT	AATATTTTCT	GTGTGAAAAA	CAACTGGTCA	AATGATGTCC
1201	GTGCTTCCCT	GTACAGTTTA	ATGGTGCTCA	TAATTCTGAC	CACACTCGTT
1251	GGCAATCTGA	TAGTTATTGT	TTCTATATCA	CACCTCAAAC	AACTTCATAC
1301	CCCAACAAAT	TGGCTCATTC	ATTCCATGGC	CACTGTGGAC	TTTCTTCTGG
1351	GGTGTCTGGT	CATGCCTTAC	AGTATGGTGA	GATCTGCTGA	GCACTGTGG
1401	TATTTTGGAG	AAGTCTTCTG	TAAAATTCAC	ACAAGCACCG	ACATTATGCT
1451	GAGCTCAGCC	TCCATTTTCC	ATTTGTCTTT	CATCTCCATT	GACCGCTACT
1501	ATGCTGTGTG	TGATCCCATC	AGATATAAAG	CCAAGATGAA	TATCTTGTTT
1551	ATTTGTGTGA	TGATCTTCAT	TAGTTGGAGT	GTCCCTGCTG	TTTTTGCATT
1601	TGGAATGATC	TTTCTGGAGC	TAAACTTCAA	AGGCGCTGAA	GAGATATATT
1651	ACAAACATGT	TCACTGCAGA	GGAGGTTGCT	CTGTCTTCTT	TAGCAAAATA
1701	TCTGGGGTAC	TGACCTTTAT	GACTTCTTTT	TATATACCTG	GATCTATTAT
1751	GTTATGTGTC	TATTACAGAA	TATATCTTAT	CGCTAAAGAA	CAGGCAAGAT
1801	TAATTAGTGA	TGCCAATCAG	AAGCTCCAAA	TTGGATTGGA	AATGAAAAAT
1851	GGAATTTTCA	AAAGAAAAGA	AAGGAAAGCT	GTGAAGACAT	TGGGGATTGT
1901	GATGGGAGTT	TTCTTAATAT	GCTGGTGCCC	TTTCTTTATC	TGTACAGTCA
1951	TGGACCCTTT	TCTTCACTAC	ATTATTCCAC	CTACTTTGAA	TGATGTATTG
2001	ATTTGGTTTG	GCTACTTGAA	CTCTACATTT	AATCCAATGG	TTTATGCATT
2051	TTTCTATCCT	TGGTTTAGAA	AAGCACTGAA	GATGATGCTG	TTTGGTAAAA
2101	TTTTCCAAAA	AGATTCAATC	AGGTGTAAAT	TATTTTGGGA	ATTGAGTTCA
2151	TAGAATTATT	ATATTTTACT	GTTTTGCAAA	TCGGTTGATG	ATCATATTTA
2201	TGAACACAAC	ATAACGAACC	ACATGCACCA	ACCACATGGA	TTTTTTTTTA
2251	AATCAGTTAC	TTGAGTCAAA	GTATGTATGG	TGAGTTAAAT	TATGATGCTT
2301	ATAGGTAATT	TCCTATTTGG	GACATAGTAG	GTATACGCTT	TTCCATTCTT
2351	ACCACACATA	ATGGAACCTT	GCAAATCCAG	TATTTAAAGG	CCTACATTTT
2401	ATATAACTTT	TCCTGCCCTT	AGAAGAACTG	CCATGAGTTT	ACTGTGGTAC
2451	CTTAGTCAGC	TGTTTCAGTG	TGGAAACTAT	AGGGCTGAAT	TTGAGGATGC
2501	AAATCAGAAT	GATTTTGCTA	TATACGTATA	CATGTCCTGG	TCACATATGT
2551	GTATAAGTCT	TTTAGTAAAA	TCAACTAAAA	GTACACACAT	TAAATGTTAC
2601	CGCTAAATAA	TTGTTGACAC	ATAATTTATT	TGTATTGTTA	ATATATATTC
2651	CAATCCTTAA	ATATCTTGAT	CTTCACATTT	TTAAATTATG	TTTTTCTGAT
2701	TCTCATGCTT	CTAAGAAATT	TGAAAAATAA	AGGAAAAATAT	AAAAATAAAA
2751	TATATACAAA	CCAAATGAAA	TTAAAAAAAT	GTTATTCATA	GTATACTTCC
2801	TGGTAAGGAT	TATATCATCT	AAAATTCCTT	ATTTTATATT	AATATTTCTC
2851	TTTTTCAACT	TTTATTTTGG	AGTCGGGGCA	TATGTGCACA	TATTAACATA
2901	ATATAATGCA	TGATGCTGAG	GTTTGAGCTA	TGAGTAATCC	CTTTTGTGG
2951	CTGTATAACG	TGACTGCACA	GATGTACAAA	TATATTTTAA	ACCATTCCCT
3001	ATTGTTAATT	CCAGTTGTTT	CTAACTTTTC	TAAAAATAAT	ATAAATTAAA
3051	TGAAATTCCCT	GCTTTTTTACC	TGCCACTGAA	GTCATGAAAA	TGTCTAGAAG
3101	GATTTTACCA	TGTCTTAAGG	TCATATCTGG	CATGATAGGG	TTCAAAACAC

FIGURE 3, sheet 1 of 2

```

3151 AGGCTACCTG GTATTAAACA AATTCAC TTT GCTGGGCTCC ACAATACACA
3201 GAAGGAGAAG CAGTCATCCA TCAAAGTAGC TAAACATGAG GGCCAACAAG
3251 AAGTATAATC AGACTAGATG TACCATGGCT ATTTAGATGG CATATATAAA
3301 AATACAAAAG AGGAACAAAT AATGGTTTCA AATAAGATTC TCTAAAGGAA
3351 GTGGGCAAAC ATTCTAAATT ACAAGCATT CAGCAATT GAGCTGATGA
3401 ATACAAACTT CACATGGTCT GTTTCAGATT GAGTATCACC GGGGATTG
3451 ATTTACCTTC TGCTCATGAA TAAGTTAGG TAAGGCAAAT GATTTTAAAC
3501 G

```

FEATURES:

Start: 1137

Exon: 1137-2153

Stop: 2151

Chromosomal map position for allelic variants:

RH Map:

11000250647308	#	SHGCNAME	CHROM#	LOD_SCORE	DIST. (cRs)
	1	SHGC-1836	6	13.79	5
	2	SHGC-12753	6	9.94	19
	3	SHGC-1877	6	9.41	13
	4	SHGC-8456	6	9.28	21
	5	SHGC-12805	6	9.22	21

SNP:

POSITION	Allele 1	Allele 2	Context
3440	c	t	Tgagctgatgaatacaaaacttcacatggctctgtttcagattgagtatcac[c/t]ggggatttggatttaccttctgctcatgaataagttagggttaaggcaaat
2525	c	t	Aactatagggctgaatttgaggatgcaaatacagaatgattttgtcatata[c/t]gtatacatgtcctggtcacatatgtgtataagtccttttagtaaaatcaac

POSITION	Allele 1	Allele 2	Protein
3440	c	t	Intron
2525	c	t	Intron

FIGURE 3, sheet 2 of 2